

2.3

Use Scatter Plots to Analyse Data

Doctors and dieticians often tell people to “eat smart.” By choosing your food carefully, you can get the nutrients that your body needs while avoiding the less healthy types of fats and carbohydrates. You can use graphs to analyse the relationship between these nutrients and the energy content of common fast foods.



20 g fat
40 g carbohydrates
95 kJ energy



15 g fat
80 g carbohydrates
125 kJ energy

Tools

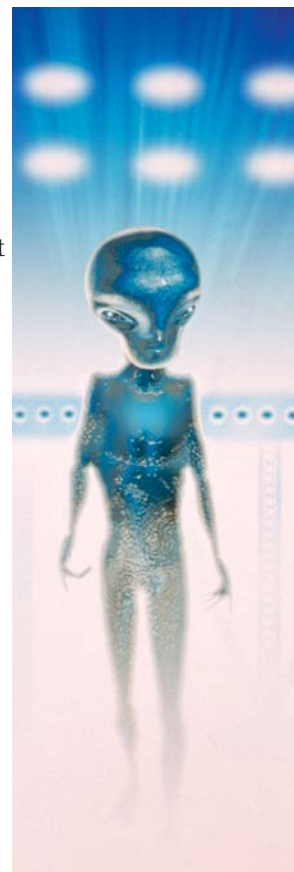
- TI-83 Plus or TI-84 graphing calculator
- grid paper

Investigate A

How can you use a scatter plot to analyse data?

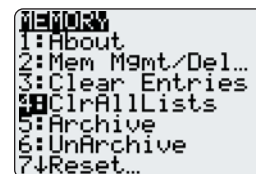
In a science fiction movie, Martians land on Earth. The scientists in the movie notice that the Martians have small hands and thin arms. The scientists measure the left forearm and left handspan of a sample of 10 Martians. The table lists these data.

Name of Martian	Left Forearm Length (cm)	Left Handspan (cm)
Enyaw	23.1	14.6
Yaj	27.3	17.2
Adnil	21.0	11.5
Sset	21.5	12.4
Nairb	26.8	15.0
Eiggam	20.2	11.1
Derf	24.4	13.2
Sirhc	23.1	15.8
Narf	24.6	17.0
Yllor	22.1	12.7



The following steps outline how to use a graphing calculator to organize and display the data.

1. First, clear any old data from the calculator's lists:
 - Press **2nd** [MEM] to display the **MEMORY** menu.
 - Move the cursor down to **4:ClrAllLists**, and press **ENTER** twice.



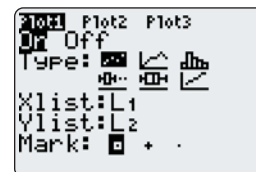
2. To start entering a table of data, press **STAT**. Then, select **1:Edit** by pressing either **1** or **ENTER**.



3. Enter the data for left forearm length into list **L1**, pressing **ENTER** after each entry. Similarly, enter the data for left handspan into list **L2**.

L1	L2	L3	1
23.1	14.6		
27.3	17.2		
21	11.5		
21.5	12.4		
26.8	15		
20.2	11.1		
24.4	13.2		
L1()=23.1			

4. Now, set the calculator to display a scatter plot of the data:
 - Press **2nd** [STATPLOT] to display the **STAT PLOTS** menu.
 - Select **1:Plot1** to display the settings for Plot 1.
 - Select **ON** if it is not already highlighted.
 - Scroll down to **Type** and select the scatter plot symbol.
 - **Xlist** should be set to **L1**. If a different list is already selected, scroll down to **Xlist:** and enter **L1**. Similarly, **Ylist** should be set to **L2**.
 - Then, press **ZOOM** and select **9:ZoomStat**.



5. What does the horizontal axis of the scatter plot represent? What does the vertical axis represent?
6. Copy the scatter plot onto a sheet of graph paper. Label your graph properly.
7. **Reflect** Describe the relationship between a Martian's left forearm length and left handspan.
8. Derf's ordered pair on the graph is (24.4, 13.2). Describe the meaning of this ordered pair. How does this ordered pair differ from the rest of the data?
9. **Reflect** What could cause the data for Derf to differ from the pattern of the other data?

Technology Tip

Pressing the **ENTER** key is like saying "Please do this." It tells the calculator to carry out the command displayed on the screen.

You can select a menu option either by pressing the number for the option or by moving the cursor to the option and pressing **ENTER**.

Literacy Connections

Your forearm is the part of your arm from your elbow to your wrist.

Your handspan is the distance from the tip of your thumb to the tip of your little finger when your fingers are spread out.

inference

- conclusion based on reasoning and data

dependent variable

- a variable that is affected by some other variable

independent variable

- a variable that affects the value of another variable

Investigate B

How do you conduct an experiment?

Is there a relationship between a human's forearm length and handspan? To answer this question, you need to collect and analyse data.

1. **Objective:** Outline the purpose of your experiment.
2. **Hypothesis:** Make a hypothesis based on the objective. Do you think a person's forearm length and handspan are related? If so, what pattern do you expect the data to have?
3. **Procedure:** Measure your own forearm length and handspan. Then, gather more data from nine classmates. Record these data in a table similar to the one for the Martian data.
4. **Observations:** Make a scatter plot of your data. Describe any pattern you see in the data.
5. **Conclusion:** Make an **inference** based on your observations. Does this inference support your hypothesis? Explain.
6. **Evaluation:** Did the experiment meet its objective? How could you improve the experiment?
7. **Reflect** How could you compare the relationship between a Martian's forearm length and handspan that you found in Investigate A to the relationship between a human's forearm length and handspan?

If you have a set of measurements of one variable, you can calculate statistics for that variable, such as the mean and the median. When you have data for two variables, you can look for a relationship between the two variables. Often, you use the data to determine whether the value of the **dependent variable** changes when the value of the **independent variable** changes.

Example 1 Identify Related Variables

Identify the independent and dependent variable in each situation.

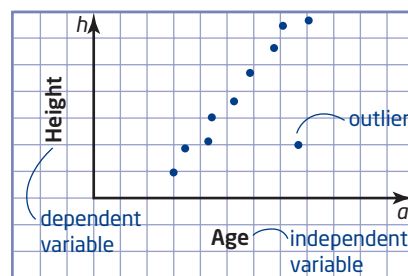
- a) Does the outdoor temperature affect the amount of fuel needed to heat a house?
- b) Is there a relationship between people's ages and their heights?
- c) Does the amount of rain in a region depend on its latitude?

Solution

- a) Since you want to know whether the outdoor temperature affects the amount of fuel required for heating, the independent variable is the outdoor temperature and the dependent variable is the amount of fuel required.
- b) The independent variable is age and the dependent variable is height.
- c) The dependent variable is the amount of rain and the independent variable is the latitude. Note that the latitude might not actually affect the amount of rain. However, to analyse the data, you treat latitude as the independent variable and the amount of rain as the dependent variable.

Often, a **scatter plot** can help you see if there is a relationship between two variables. On such graphs, the horizontal axis usually represents the independent variable, while the vertical axis represents the dependent variable. A measurement of the independent variable and the corresponding measurement of the dependent variable make up an ordered pair, (x, y) , which appears as a point on the scatter plot.

An **outlier** is a point separated from the main body of data on a graph. Sometimes, an outlier results from a measurement error or from some factor that affects only a few of the observed values for a variable. If you can show that an outlier is inaccurate or unrepresentative, you can leave it out of your calculations. Otherwise, you should include the outlier in the data set.



outlier

- measurement that differs significantly from the rest of the data

Example 2 Draw a Scatter Plot

A skateboarder starts from various points along a steep ramp and practises coasting to the bottom. This table lists the skateboarder's initial height above the bottom of the ramp and his speed at the bottom of the ramp.

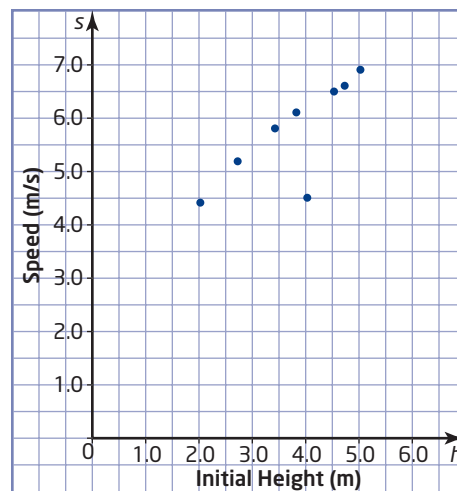
Initial Height (m)	2.0	2.7	3.4	3.8	4.0	4.5	4.7	5.0
Speed (m/s)	4.4	5.2	5.8	6.1	4.5	6.5	6.6	6.9

- Identify the independent variable and the dependent variable. Explain your reasoning.
- Make a scatter plot of the data.
- Describe the relationship between the variables.
- Identify any outliers. What might cause an outlier in the data?

Solution

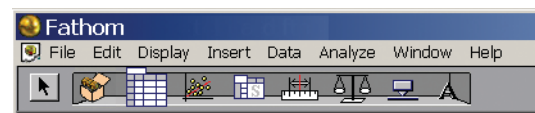
- The skateboarder's speed at the bottom of the ramp depends on how high up the ramp he starts. So, the independent variable is the initial height, and the dependent variable is the skateboarder's speed at the bottom of the ramp.

b) Method 1: Use Pencil and Paper



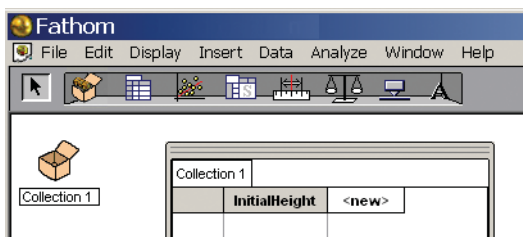
Method 2: Use *Fathom™*

Click on the case table icon and drag it onto the desktop.



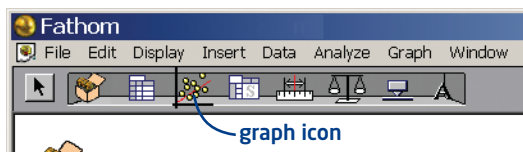
case table icon

Click on <new> in the first column, and enter “InitialHeight.” Click on <new> in the second column, and enter “Speed.”

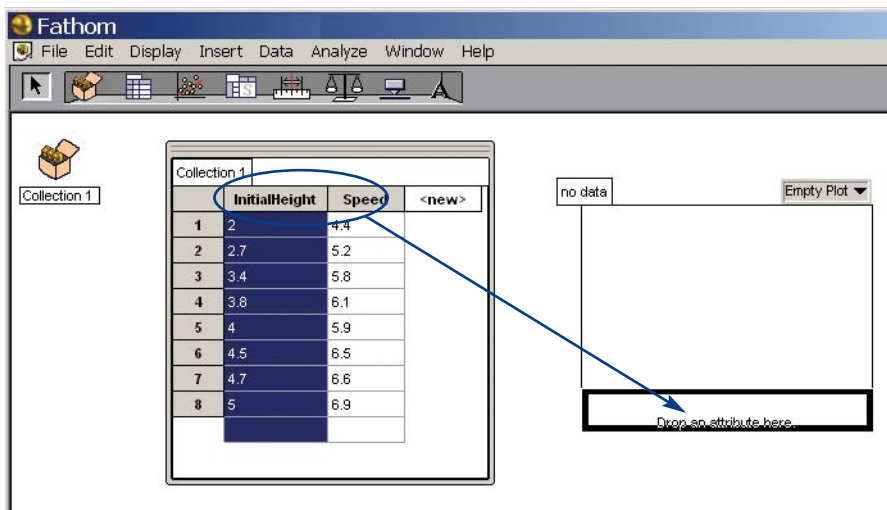


Enter the data in the case table.

Click on the graph icon and drag it onto the desktop.



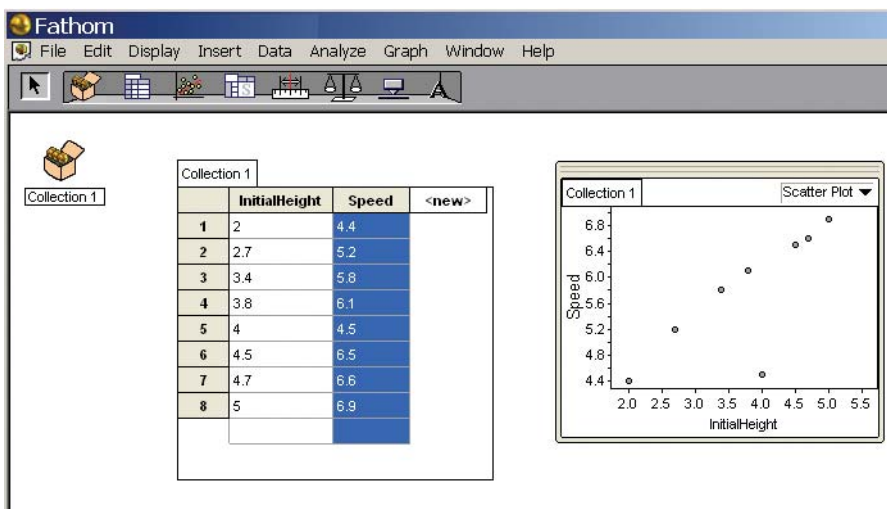
Drag the “InitialHeight” heading to the horizontal axis.



Technology Tip

You can use **Ctrl-Z** to undo an action. This shortcut works in many programs.

Drag the “Speed” heading to the vertical axis. *Fathom*™ then generates the scatter plot automatically.



Technology Tip

You can delete the legend box from your graph.

In Quattro® Pro, click on the graph, then click **Chart** on the toolbar. Select **Legend**, click on the image of a graph without a legend, and click OK.

In Excel, right-click on the graph, select Chart Options, click the **Legend** tab, and uncheck the **Show Legend** box.

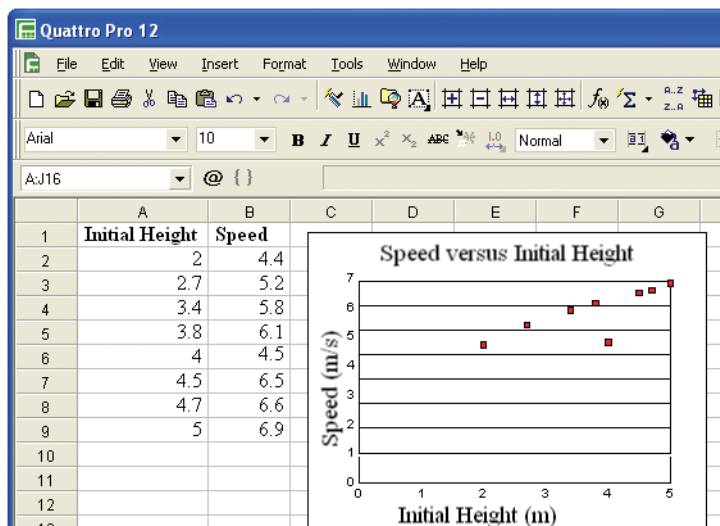
Method 3: Use a Spreadsheet

Enter the data in the first two columns of the spreadsheet. Then, select these data.

In Corel® Quattro® Pro, click **Insert/Chart**. Then, click **Next**.

Uncheck the **3D** box. For chart type, click on **Scatter**, and select the **no line** option. Click **Next** again.

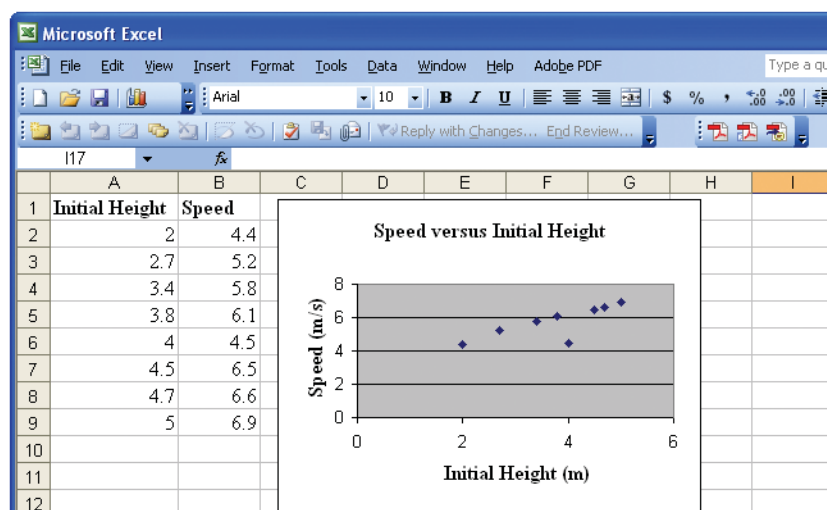
Enter the title for your graph and the labels for the axes. Click **Finish**. Then, move the cursor to where you want the graph to appear, and click to place it.



In Microsoft® Excel, click **Insert/Chart**. Under Chart type, click on **XY (Scatter)**, and click the **Next** button twice.

Enter the title for your graph and the labels for the axes.

Click the **Next** button again, and then click **Finish**.



You can make your graph easier to read by enlarging it. Click on the graph, then drag a side or corner.

You can also adjust the scale of the axes.

- In Quattro® Pro, click on the axis to select it, then right-click. Select **Axis Properties**, click on the **Scale** tab, and enter the settings you want.
 - In Excel, right-click on the axis and select **Format Axis**. Then, click on the **Scale** tab and enter the settings you want.
- c) As the initial height increases, the skateboarder's speed at the bottom of the ramp increases. The pattern of the data points is a curve rather than a straight line.
- d) The ordered pair (4.0, 4.5) is an outlier because it is separated from the rest of the data. The speed is less than you would expect from looking at the rest of the data. The skateboarder could have dragged his foot or started at an angle that slowed him down.

Key Concepts

- Scatter plots can help you see relationships between variables.
- Graphing calculators, spreadsheets, and statistics software can plot graphs of data.
- When the value of the independent variable changes, the value of the dependent variable may also change.
- An outlier is separate from the main body of data. You can discard an outlier only if you know that it not representative of the relationship between the variables.

Communicate Your Understanding

- C1** Explain how a scatter plot can show that two variables are related.
- C2** Which statement is true? Why?
- A** You should discard all outliers since they do not follow the pattern of the rest of the data.
 - B** You should not discard an outlier unless you can show that it does not accurately represent the relationship between the variables.

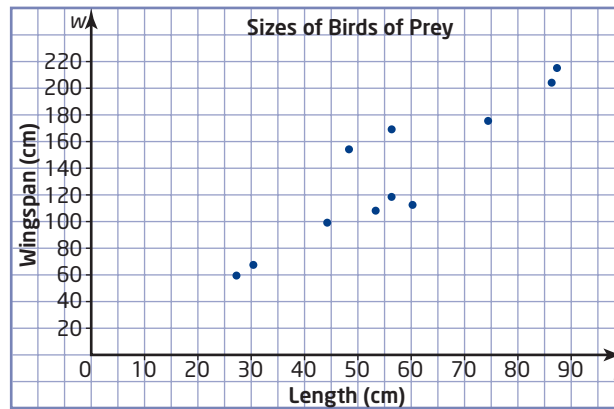
Practise

For help with question 1, see Example 1.

1. Identify the independent and the dependent variable in each pair.
 - a) blood pressure and physical fitness
 - b) income and level of education
 - c) load in an airplane and length of runway needed for takeoff

For help with questions 2 and 3, see Example 2.

2. This scatter plot compares the body lengths and wingspans of 11 birds of prey.



- a) How would you change the graph to show wingspan as the independent variable?
 - b) Describe the relationship between length and wingspan of the birds of prey.

3. This table shows the numbers of days absent from science class and the report card marks for 15 students.

- a) Identify the independent variable and the dependent variable. Explain your reasoning.
 - b) Make a scatter plot of the data.
 - c) Describe the relationship between a student's marks and attendance.
 - d) Are there any outliers? If so, explain how they differ from the rest of the data.

Number of Days Absent	Mark (%)
2	81
0	73
12	50
7	63
1	77
22	38
10	44
3	95
3	56
4	71
8	67
0	78
9	61
15	40



Connect and Apply

In these questions, you can draw the scatter plots by hand or produce them with a graphing calculator, Fathom™, or a spreadsheet.

4. This table lists the data from an experiment to measure how high a ball bounces after being dropped from six different heights.

Initial Height (m)	Bounce Height (m)
1.00	0.62
1.50	0.91
2.00	1.18
2.50	1.51
3.00	1.80
3.50	2.08

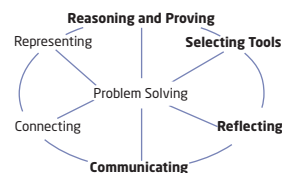
- a) Identify the independent variable and the dependent variable. Explain your reasoning.
- b) Make a scatter plot of the data.
- c) Describe the relationship between the initial height of the ball and its bounce height.
- d) If the data included the ordered pair (4.00, 1.62), would you consider it to be an outlier? Would you discard this ordered pair from the data set? Explain your reasoning.
5. This table shows the distance it took a car to stop when travelling at various speeds.

Speed (km/h)	40	50	60	70	80	90	100	120	140	160
Stopping Distance (m)	12	14	21	35	41	51	62	84	122	159

- a) Make a scatter plot of these data.
- b) Describe the relationship between the speed of a car and its stopping distance.
- c) An additional measurement produces the ordered pair (85, 46). Explain the meaning of this pair. Is it an outlier? Explain why or why not.
6. a) Make a hypothesis about the relationship between a person's height and shoulder width.
- b) Design and carry out an experiment to investigate the relationship. What conclusions can you make from the data you collected?
- c) Compare your hypothesis with the results of your experiment.
- d) How could you improve your experiment?

Literacy Connections

Shoulder width is the distance from the outside of one shoulder to the outside of the other.



7. Chapter Problem In Section 2.1, question 8, you made a hypothesis about a physical characteristic that could help an athlete do well in the high jump.

- Outline an experiment to examine the relationship between this physical characteristic and the height a person can jump. Describe how you could collect the data.
- Identify the independent and the dependent variable in your experiment.
- Describe how you could use a scatter plot to analyse the data. What pattern would you see in the scatter plot if your hypothesis were true?

8. This table shows the fat and energy content in typical servings of fast food.

Item	Serving Size (g)	Fat (g)	Energy (kJ)
Harvey's Original Hamburger	157	20	410
Harvey's Veggie Burger	142	9	317
Mr. Submarine Small Assorted Sub	179	6	280
Mr. Submarine Small Vegetarian Sub	117	3	180
Pizza Pizza Pepperoni Slice (walk-in)	277	19	630
Pizza Pizza Vegetarian Slice (walk-in)	326	14	580
KFC Chicken Breast	161	19	380
KFC Popcorn Chicken	114	21	380
Swiss Chalet Quarter Chicken Breast	159	12	300
Swiss Chalet Garden Salad (no dressing)	162	0	30
Swiss Chalet Caesar Salad	170	32	360

- Calculate the amount of fat, in milligrams, per gram of each item. Round to the nearest milligram. Then, calculate the energy content per gram of each item. List the results of your calculations in a table.
- Make a scatter plot of the two sets of data you calculated in part a). Would you put the fat per gram or the energy per gram on the horizontal axis? Why?
- Identify and explain any outliers. Should they be removed from the data set? Explain.
- Describe what you can learn from the scatter plot.

Extend

9. The costume designers for a science-fiction movie decide to make the Martians' heights about 6 times their neck circumferences. North Americans' heights are about 4.5 times their neck circumferences. Draw a graph of both relationships on the same grid.
10. This table lists the number of times at bat and the numbers of doubles, home runs, and strikeouts for starting first basemen in American League baseball during the 2004 season.

Player	At Bats	Doubles	Home Runs	Strikeouts
B. Broussard	418	28	17	95
C. Delgado	458	26	32	115
D. Erstad	495	29	7	74
J. Giambi	264	9	12	62
K. Harvey	456	20	13	89
S. Hatteberg	550	30	15	48
P. Konerko	563	22	41	107
T. Martinez	458	20	23	72
D. Mientkiewicz	391	24	6	56
J. Morneau	280	17	19	54
J. Olerud	425	20	9	61
R. Palmero	550	29	23	61
C. Pena	481	22	27	146
M. Sweeney	411	23	22	44
M. Teixeira	545	34	38	117

Did You Know?

The earliest known baseball game in Canada took place in Beachville, Ontario on June 14, 1838.

- a) Make a scatter plot of doubles per at bat versus home runs per at bat.
 - b) Make a scatter plot of doubles per at bat versus strikeouts per at bat.
 - c) Make a scatter plot of home runs per at bat versus strikeouts per at bat.
 - d) Do any of the scatter plots show a relationship between the variables?
11. **Math Contest** Each of the letters in the expression $\frac{a}{b} + \frac{c}{d} + \frac{e}{f}$ represents a different integer from 1 to 6 inclusive. Find the smallest possible value for $\frac{a}{b} + \frac{c}{d} + \frac{e}{f}$.